

Update on SOAR 9 and Sentence Processing

Deryle Lonsdale,
Jeremiah McGhee,
Ross Hendrickson
and the BYU SOAR Research Group

The Headlines

- NLSOAR is in active development
- NLSOAR can process several key types of sentences
- NLSOAR updated to SOAR 9 as a prototype
 - including the Tcl interface
 - MANY THANKS to Bob Mariner and Jon Voigt for lots of quality and timely help

The not so great news

- NLSOAR is not yet ready for prime-time
- But
- Presenting portion of work at ICCM
 - Lexical Access
- Good trajectory for continued development
 - 3 graduate students
 - 3 under-graduate students

Broad Strokes - Purpose

- Thorough testing of Minimalist Program
 - Something not done before for this theory
- Several approaches attempted
 - In an effort to maintain strict MP assumptions
 - -Adger, Chomsky
- Hierarchy of Projections, Merge, Adjunction
 - All implemented in prototype

Need

- *“How seriously should we take Minimalist Syntax?”*
 - Shimon Edelman and Morten H. Christiansen
TRENDS in Cognitive Sciences Vol.7 No.2 2003
- *“The minimalist program in syntax.”*
 - Howard Lasnik
TRENDS in Cognitive Sciences Vol. 6 No.4 2002
- Gist: Other approaches have been tested, Minimalism needs to face the same experimentation.



Three Key Sentence Types

- Transitivity
 - Di-transitive
 - The Giraffes deem Hyenas dross.
 - Transitive
 - Tigers chase giraffes.
 - In-transitive
 - Dogs bark.

WordNet

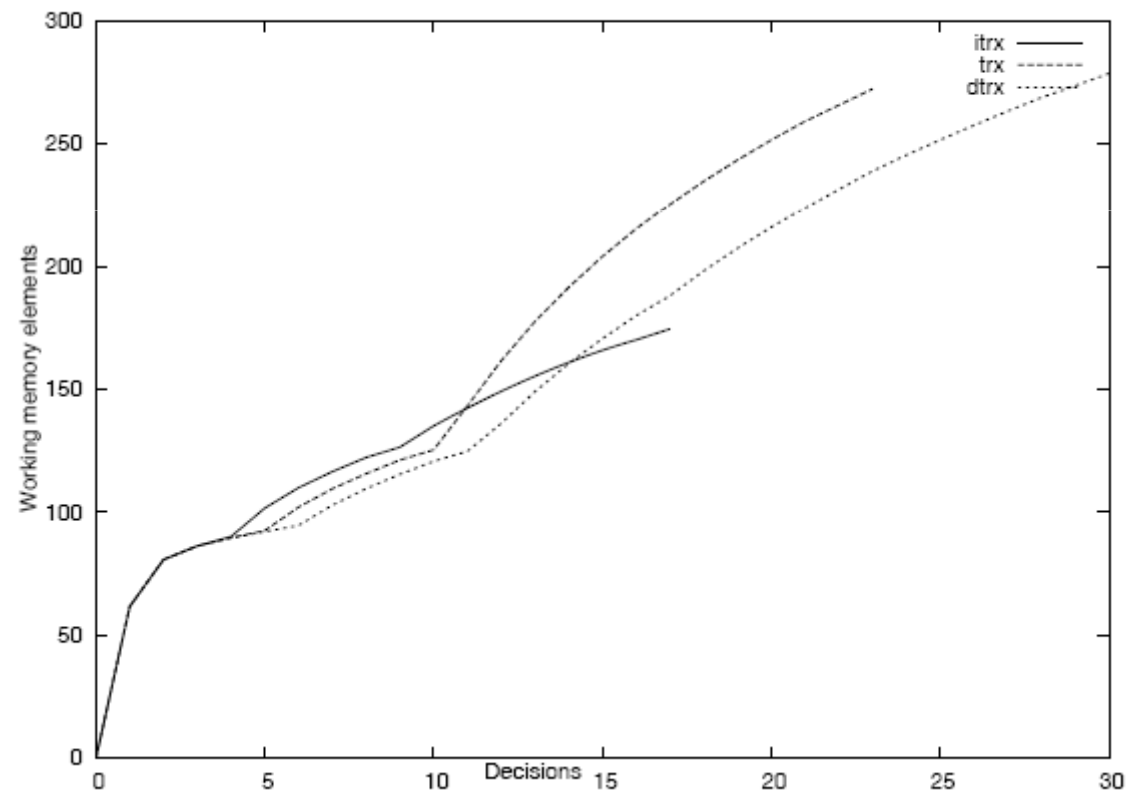
- Rewrote lexical access module in PERL
- SOAR able to use sentence frames, and sense information
 - Classify verbs as itr, tr, ditr
 - Use frame information to help build structure

WordNet

- Cost to access is high
 - Cognitively plausible?
- Working on optimizing PERL to further speed up lexical access
 - Possibly bring access into SOAR itself
- Limited Lexicon
 - Determiners, Prepositions, and Modals all require work-around for implementation
 - Determiners already implemented

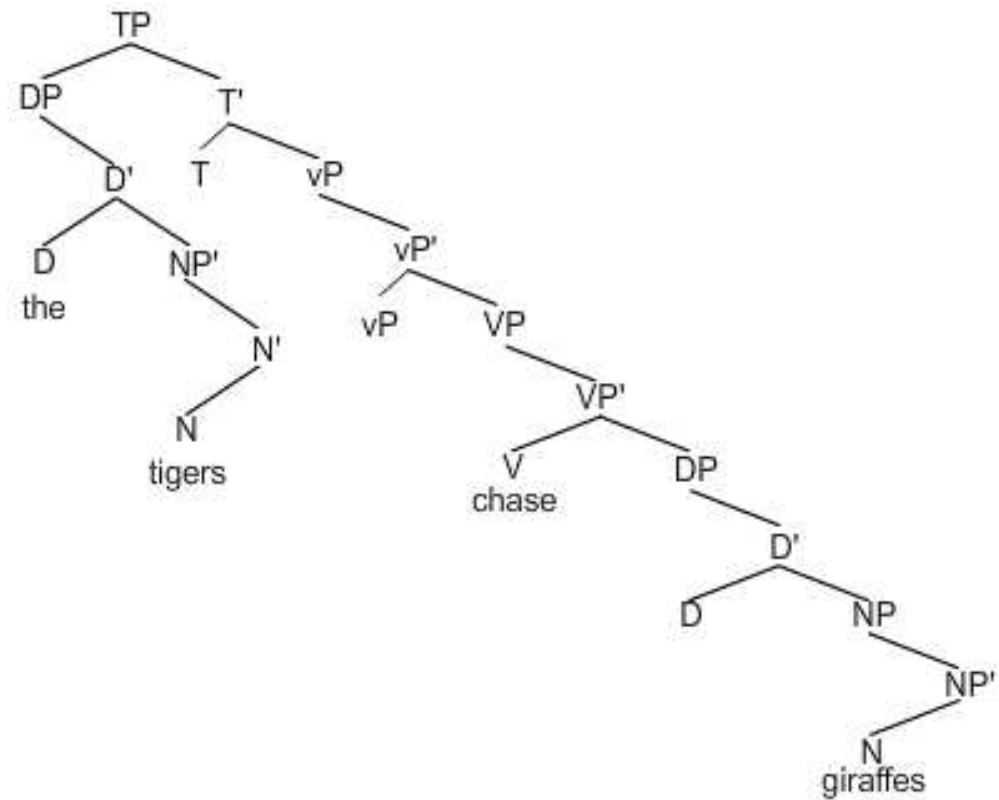
WordNet

- Cost of Lexical Access



Clig

- CLIG
 - Command to file
 - MP tree output





Learning - Chunking - A/R Set

- Not implemented in current prototype
- In process of abstracting away from structure specific operators to more generalized
- Next big step to ready NLSOAR for real world

More to do

- Corpus of 3 types of transitive constructions extracted from Brown and Wall Street Journal corpora
 - Unusable for the most part until Lexicon additions to WordNet completed
 - Large amount of exemplars of real world usage to put through SOAR
- Add distinction in Intransitive
 - Unergative
 - Unaccusative

Other Remaining Work

- Port remaining SOAR7 code:
 - learning
 - LCS semantics (which has never yet been done w/ the MP)
 - language generation
 - rudimentary discourse moves via discourse recipes
 - language translation
 - language learning
 - speech recognition front-end
 - etc

Conclusion

- Coals
 - Continued problems integrating code (C, TCL)
 - Keeping up to date:
 - Reinforcement, Episodic, Semantic Memory Modules
 - Support resources are limited
- Nuggets
 - Functional in SOAR 9
 - Still receiving expressions of interest from range of sources
 - Development team is growing
 - *Incremental processing and resource usage (ICCM)*